REMARKS

This is in response to the Office Action dated February 1, 2007. Claims 1 to 4 are unamended. New claim 5 is generally patterned after claim 1 but is amplified to even more clearly distinguish over the cited prior art.

In the Office Action, the claims were rejected under 35 U.S.C. § 103(a) as being unpatentable over the Kargupta patent in view of the Kamath patent and further in view of the Cho article.

First, the undersigned would like to thank examiner Daye and her supervisor Sana Al-Hashemi for the courtesies extended during an interview on April 26, 2007. In attendance were the undersigned, the applicant, Jerzy Bala, and Ali Hadjarian, an employee of applicant's assignee. At the interview, Drs. Bala and Hadjarian explained how the prior art fails to disclose the limitations in claim 1, utilizing the enclosed Power Point. The examiners seem to agree that the cited prior art did not disclose "beginning attribute selection" and "selecting a winning agent," and said they would consider those limitations further. They disagreed with many of the other positions advanced by Drs. Bala and Hadjarian.

The Office Action sets forth how each limitation of the claims is met by certain ones of the three items of prior art. We respectfully disagree as set forth below. The reasons are set forth below under headings corresponding to the various limitations in claim 5.

Claim limitation: beginning attribute selection by each agent, wherein attribute selection of one data attribute from a set of local data attributes unique to the respective agent such that the selected data attribute has the substantially highest local information gain value among all attributes

The Office Action states that Kargupta discloses beginning attribute selection, citing col. 3, lines 20-27, which states:

> "Given a set of observed feature values, the task is to learn a function that computes the unknown value of a desired feature as a function of other observed features. The given set of observed feature values is sometimes called the training data set. In FIG. 1 the col. for *f* denotes the feature value to be predicted; x_1, x_2, x_3, x_4, x_5, x_6 and x_7 denote the features that are used to predict *f*."

Here Kargupta is pointing to the estimation of an unknown feature value based on the value of the known features. This does not constitute "attribute selection by each agent." In other words, there is no equivalency between attribute "estimation" (i.e., computing the unknown value of an attribute) and attribute "selection" (i.e. picking the best attribute for the classification task at hand).

We submit that attribute/feature selection can be considered as a pre-processing step prior to function value estimation/prediction. In other words, the objective of attribute selection is to find the most meaningful attribute for function value estimation/prediction. An example of function value estimation/prediction is predicting a patient's cancer risk (function) based on the attributes such as sex, smoking habit, age and salary. An example of attribute selection is the precursor step of deciding that smoking habit is the most discriminatory attribute for predicting a patient's cancer risk as it has the highest information gain value.

Claim limitation: collecting the highest information gain values from the plurality of agents by the mediator, wherein the highest information gain value of a respective agent is based on its own local data with its own unique data attributes

The Office Action states that Kargupta discloses passing a best attribute, citing col. 13, lines 18-27, which states:

> "We can express Boolean decision tree as a function $f:X^n \rightarrow \{0, 1\}$. The function *f* maps positive and negative instances to one and zero respectively. A node in a tree is labeled with a feature x_i. A downward link from the node x_i, is labeled with an attribute value of i-th feature. A path from the root node to a successor node represents the subset of data that satisfies the different feature values labeled along the path. These data subsets are essentially similarity based equivalence classes and we shall call them schemata (schema in singular form). If h is a

> schema, then $h_e\{0, 1, *\}^1$, where * denotes a wild-card that matches any value of the corresponding feature."

Here, Kargupta is providing the reader with a high level description of the ID3 algorithm. ID3, developed by Ross Quinlan, is arguably the most popular algorithm for decision tree construction. The whole motivation behind the invention of claim 5 has been to come up with a distributed version of the ID3 algorithm.

Kargupta does not disclose passing a best attribute from each of said plurality of agents to said mediator. The patent merely describes the standard centralized approach to decision tree construction.

We submit that Kargupta discloses the centralized approach to decision tree construction where there are only global attributes and globally optimum information gain values, whereas the invention of claim 5 involves a plurality of agents each with its own local attributes and locally optimum information gain values. The information on each agent's highest information gain value is collected by the mediator for comparison purposes.

Claim limitation: selecting by the mediator of a winning agent, wherein the winning agent is the only agent from the plurality of agents with access to the local data attribute with the highest global information gain value

The Office Action states that Kamath discloses selecting a winning agent from said plurality of agents, citing col. 14, lines 9-19, which states:

> "Each processor evaluates each of the local feature lists to find the best local split (this is done in parallel by all processors).
>
> "It communicates the local best splits and count statistics to all processors.
>
> "Each processor determines the best global split (this is done in parallel by all processors).
>
> "Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors."

Here, Kamath is not disclosing selecting a winning agent from said plurality of agents. He is merely explaining the process of selecting a winning feature, which "*is done in parallel by all processors.*" (emphasis added) This is a significant algorithmic difference.

We submit that in Kamath's approach, all processors (agents) have access to the winning feature (attribute). In the invention of claim 5, only a single agent has access to the winning attribute.

Claim limitation: initiating data splitting by said winning agent based on the value of the data attribute with the highest information gain wherein the specified data attribute is unique to the respective agent's local data

The Office Action states that Kamath discloses initiating data splitting, citing col. 13, lines 56-60 by a winning agent, citing col. 14, lines 9-19. The cited passages are:

> "The creation of the tree is thus split into two parts:
>
> "(1) Initial Sorting
>
> "First the training set is split into separate feature lists for each feature. Each list contains the identification (ID) number of the data instance, the feature value, and the class associated with the instance. This data is partitioned uniformly among the processors."
>
> "Each processor evaluates each of the local feature lists to find the best local split (this is done in parallel by all processors).
>
> "It communicates the local best splits and count statistics to all processors.
>
> "Each processor determines the best global split (this is done in parallel by all processors).
>
> "Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors."

Kamath does not disclose data splitting by the winning agent. Instead, the patent discloses splitting by all agents (or processors) based on a winning feature. According to Kamath, this is a process that is done in parallel by all processors. This is algorithmically quite

different from the invention of claim 5 where initial splitting is only performed by the winning agent (i.e. the only agent with access to the data attribute with the highest information gain). Since non-winning agents do not have access to the corresponding feature, data splitting by them will not be feasible.

Claim limitation: forwarding split data index information resulting from said data splitting by said winning agent to said mediator

The Office Action states that Kamath discloses forwarding split data index information resulting from said data splitting by said winning agent to said mediator, citing the above-quoted text in col. 14, lines 9-19.

We submit that Kamath does not disclose forwarding split data index information resulting from data splitting by the winning agent. Instead, Kamath discloses the forwarding of all split data index information by all the agents (or processors). This is a significant algorithmic difference.

Claim limitation: initiating data splitting by each of said plurality of agents other than said winning agent based on the split data index information furnished by the winning agent and broadcasted by the mediator

The Office Action states that Kamath discloses initiating data splitting by each of said plurality of agents other than said winning agent, citing col. 14, lines 17-26, which states:

> "Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors.
>
> "Then, each processor builds a hash table containing all the ID numbers, and information on which instances belong to which decision tree node.
>
> "Next, each processor, for each feature, probes the hash table for each ID number to determine how to split that feature value.
>
> "This process is carried out on the next unsolved decision tree."

Again, Kamath does not disclose initiating data splitting by each of said plurality of agents "other than said winning agent." In fact, Kamath clearly states that data splitting is performed by all the agents (or processors). As stated previously we submit that such splitting by all processors is not feasible in the invention of claim 5, as only one agent has access to the data attribute with the highest information gain.

Claim limitation: generating and saving partial rules by repeating the attribute selection and data splitting process recursively and by tracking the attribute/split information coming from that iteration's winning agent

The examiner states that Cho discloses generating and saving partial rules, citing p. 2, lines 14-18, which states:

> "Besides devising a new technique named 'fragmentation approach' in this paper, we also investigate various technical details, such as how large a data set in each fragment (local data set) is adequate, how many rules are to be generated in each fragment, and the corresponding selection criterion by which to choose the generated rules from the fragments to form a global rule set."

Cho does not disclose generating and saving partial rules. Instead, the article is referring to the process of generating whole rules at each fragment and then using them to form a global rule set. In contrast, the invention of claim 5 does not advocate constructing whole rules at each agent location. Instead, the invention of claim 5 constructs rule "conditions" at each location.

For example consider the following two rules:

A & B -> C1 (Rule 1)

C & D -> C3 (Rule 2)

Cho deals with generating all of Rule 1 at one fragment and all of Rule 2 at another and then combining them to get the rule set.

Instead, the invention of claim 5 would generate A at one location, and B at another location, so that combining them through a mediator would finally reveal the whole rule: A & B -> C1.

Claim limitation: outputting complete rules obtained at the completion of the mining process to said plurality of agents

The examiner states that Cho discloses outputting complete rules to said plurality of agents, citing p. 4, lines 24-25, which states:

> "Another strategy is to loosely account on an number of different inductive learning algorithms by integrating their collective output concepts."

Cho does not disclose outputting complete rules to said plurality of agents. Instead, the article is referring to the integration of the outputs of a number of different learning algorithms. This deals with the topic of "multiple classifier" approaches, which is a whole discipline unto itself and does not have anything to do with the invention of claim 5.

Accordingly, it is submitted that a number of limitations of claim 5 are not met by the cited prior art. Accordingly, claim 5 is patentable over the cited art. Claims 1-4 include generally the same limitations as claim 5, and are therefore patentable for the same reasons. It is therefore submitted that this application is in condition for allowance and such action is requested.

Respectfully submitted,
Harold V. Stotland
Reg. No. 24,492
Seyfarth Shaw
Attorneys for Assignee
55 East Monroe Street
Suite 4200
Chicago, Illinois 60603-5803
312-346-8000

By

CH1 11226093.1

Application No. **10/616,718**
vs.
Kamath
Kargupta
Cho

Presented to

United States Patent and Trademark Office

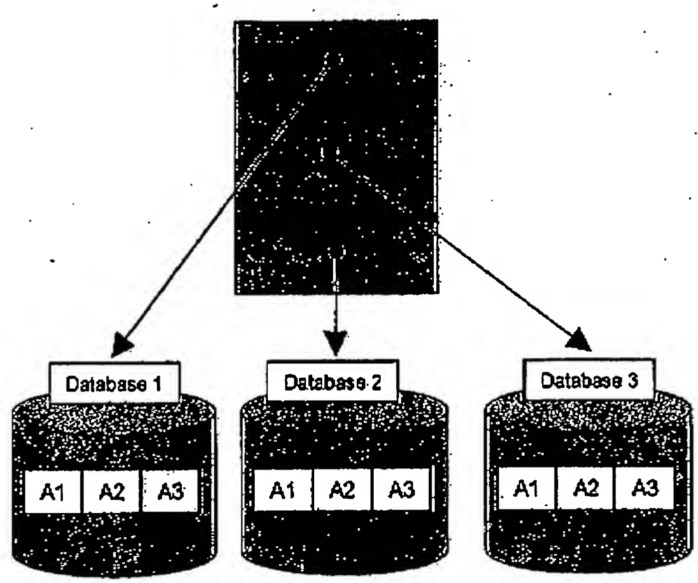
April 26, 2007

Major Differences

- Distributed Data Type
  - heterogeneous vs. homogeneous
  - parallel vs. distributed/sequential computing
- Winning Feature Selection and Data Split
  - global vs. local

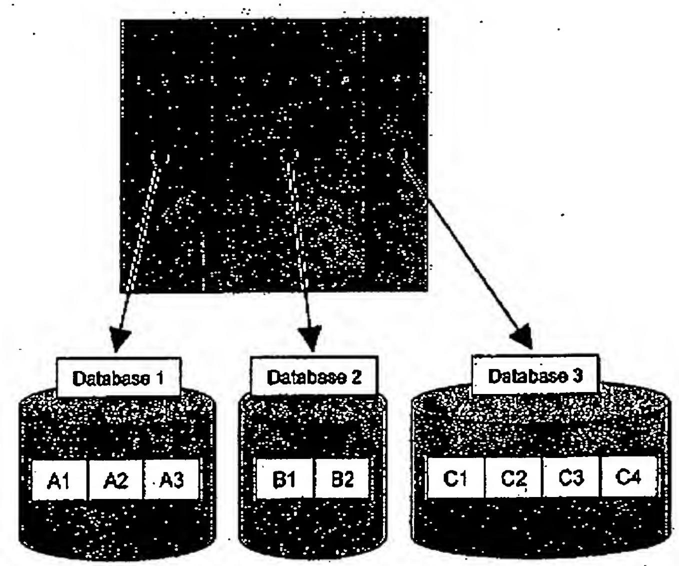
Distributed Data Type

Horizontal Data Partitioning
(homogenous)



Parallel computing

Vertical Data Partitioning
(heterogeneous)



Distributed/sequential computing

Winning Feature Selection and Data Splits

Kamath

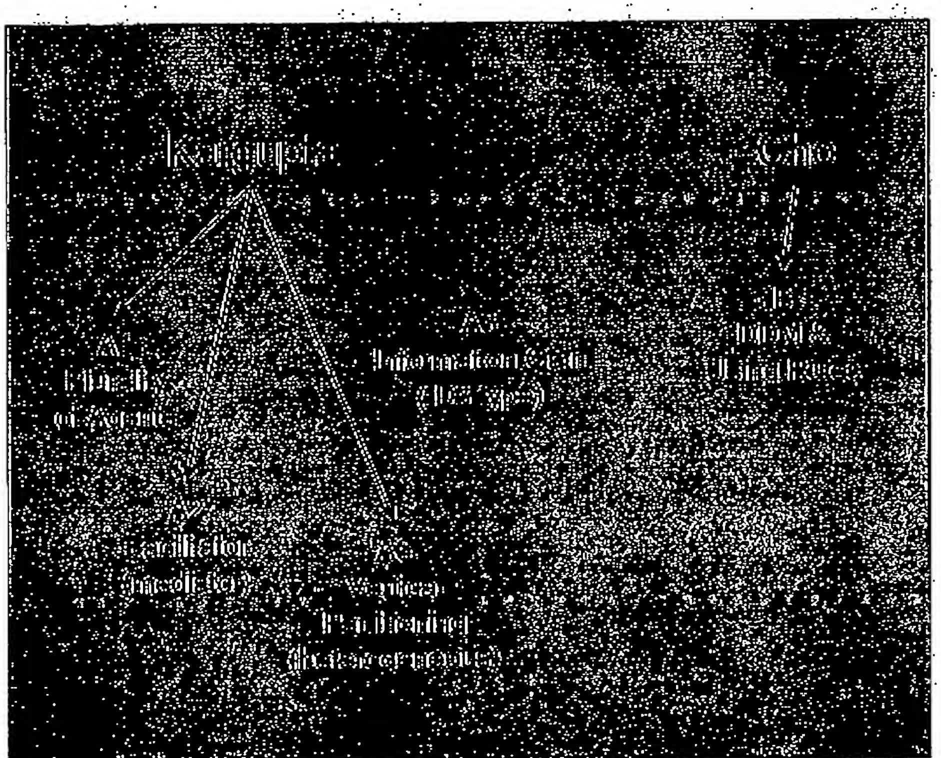
Each processor evaluates each of the local feature lists to find the best local split (this is done in parallel by all processors).
It communicates the local best splits and count statistics to all processors.
Each processor determines the best global split (this is done in parallel by all processors).
Split the Data. <u>Each processor splits</u> on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors.
Then, each processor builds a hash table containing all the ID numbers, and information on which instances belong to which decision tree node.
Next, each processor, for each feature, probes the hash table for each ID number to determine how to split that feature value.

Can InferAgent's approach be inferred from Kamath ?

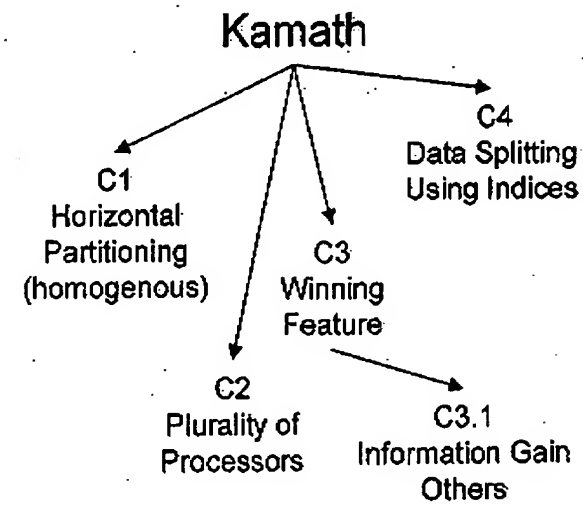
InferAgent

Each Agent evaluates each of the local feature lists to find the best local split.
It communicates the local best splits <u>TO MEDIATOR</u>.
MEDIATOR DETERMINES THE BEST GLOBAL SPLIT BY COMPARING CONTRIBUTED BEST LOCAL SPLITS, COMMUNICATES THIS TO THE WINING AGENT (i.e., THE ONE WITH THE BEST GLOBAL SPLIT MEASURE), AND INSTRUCTS THIS AGENT TO SPLIT THE DATA.
<u>ONLY WINING AGENT SPLITS</u> THE DATA AND SENDS THE ID (INDICES) OF ITS NEW LEFT AND RIGHT NODE DATA INSTANCES TO OTHER AGENTS.
EACH AGENT SPLITS DATA BASED ON INDICES OF THE LEFT AND RIGHT NODE DATA INSTANCES THAT WERE DETERMINED BY THE WINING AGENT (THE BEST GLOBAL SPLIT).

Major Components in Three References



Syntactic Correspondence to InferAgent's Approach
(different algorithmic methods)



CLAIM LIMITATION A

Claim Limitation	Examiner's Position	Cited Language	Bala Position
invoking agents by a mediator	Kargupta discloses: invoking agents by a mediator (Fig.8; column 3, lines 60-66, Kargupta; wherein facilitator corresponds with mediator);	Kargupta (column 3, lines 60-66) states: *An agent coordinating facilitator gives user requests to local agents which then access and analyze local data, returning analysis results to the facilitator, which merges the results.*	Giving "user requests" to local agents, in Bala's opinion, is not equivalent to agents being invoked by a mediator. In Bala invention, there are no user requests. Furthermore, in Bala invention, unlike Kargupta's, there is no passing of information from the mediator to the agents at this initial step of being "invoked".

CLAIM LIMITATION B

Claim Limitation	Examiner's Position	Cited Language	Bala Position
beginning attribute selection by a plurality of agents	Kargupta discloses: beginning attribute selection (column 3, lines 20-27, Kargupta) by a plurality of agents (Fig.8, Kargupta)	Kargupta (column 3, lines 20-27) states: *Given a set of observed feature values, the task is to learn a function that computes the unknown value of a desired feature as a function of other observed features. The given set of observed feature values is sometimes called the training data set. In FIG. 1 the column for .function. denotes the feature value to be predicted; x.sub.1, x.sub.2, x.sub.3, x.sub.4, x.sub.5, x.sub.6 and x.sub.7 denote the features that are used to predict .function..*	Here Kargupta is pointing to the estimation of an unknown feature value based on the value of the known features. This by no means constitutes "attribute selection by a plurality of agents". In other words, there is no equivalency between attribute "estimation" (i.e., computing the unknown value of an attribute) and attribute "selection" (i.e. picking the best attribute for the classification task at hand).

CLAIM LIMITATION C

Claim Limitation	Examiner's Position	Cited Language	Bala Position
passing a best attribute from each of said plurality of agents to said mediator wherein a best attribute is an attribute having a highest information gain as between attributes found by the respective agent	Kargupta discloses: passing a best attribute (column 13, lines 18-27, Kargupta) from each of said plurality of agents to said mediator (column 28, lines 49-55, Kargupta) wherein a best attribute is an attribute having a highest information gain as between attributes found by the respective agent (column 13, lines 36-58, Kargupta)	Kargupta (column 13, lines 18-27) states: *We can express boolean decision tree as a function .function..X.sup.n.fwdarw.{0, 1}. The function .function. maps positive and negative instances to one and zero respectively. A node in a tree is labeled with a feature x.sub.i. A downward link from the node x.sub.i is labeled with an attribute value of i-th feature. A path from the root node to a successor node represents the subset of data that satisfies the different feature values labeled along the path. These data subsets are essentially similarity based equivalence classes and we shall call them schemata (schema in singular form). If h is a schema, then h.epsilon.{0, 1, *}.sup.l, where * denotes a wild-card that matches any value of the corresponding feature.*	Here, Kargupta is providing the reader with a high level description of the ID3 algorithm. ID3, developed by Ross Quinlan, is arguably the most popular algorithm for decision tree construction. The whole motivation behind Bala's invention has been to come up with a distributed version of the ID3 algorithm. So in Bala's opinion, Kargupta does not disclose passing a best attribute (column 13, lines 18-27, Kargupta) from each of said plurality of agents to said mediator. He merely explains the standard centralized approach to decision tree construction.

CLAIM LIMITATION D

Claim Limitation	Examiner's Position	Cited Language	Bala Position
selecting a winning agent from said plurality of agents by said mediator.	Kamath discloses: selecting a winning agent from said plurality of agents (column 14, lines 9-19, Kamath). The combination of Kargupta in view of Kamath, disclose: the mediator selecting the winning agent (column 29, lines 1-20, Kargupta);	Kamath (column 14, lines 9-19) states: *Each processor evaluates each of the local feature lists to find the best local split (this is done in parallel by all processors).* *It communicates the local best splits and count statistics to all processors.* *Each processor determines the best global split (this is done in parallel by all processors).* *Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors.*	Here, Kamath is not disclosing selecting a winning agent from said plurality of agents. He is merely explaining the process of selecting a winning feature, which "*is done in parallel by all processors*". This is a significant algorithmic difference.

CLAIM LIMITATION E

Claim Limitation	Examiner's Position	Cited Language	Bala Position
initiating data splitting by said winning agent	The combination of Kargupta in view of Kamath, disclose: Initiating data splitting (column 13, lines 56-60, Kamath) by said winning agent (column 14, lines 9-19, Kamath)	Kamath (column 13, lines 56-60) states: *The creation of the tree is thus split into two parts:* *(1) Initial Sorting* *First the training set is split into separate feature lists for each feature. Each list contains the identification (ID) number of the data instance, the feature value, and the class associated with the instance. This data is partitioned uniformly among the processors.* Kamath (column 14, lines 9-19) states: *Each processor evaluates each of the local feature lists to find the best local split (this is done in parallel by all processors).* *It communicates the local best splits and count statistics to all processors.* *Each processor determines the best global split (this is done in parallel by all processors).* *Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors.*	Again, Kamath does not disclose data splitting by said winning agent. Instead, he discloses splitting by all agents (or processors) based on a winning feature. According to him, this is a process that is done in parallel by all processors. This is algorithmically quite different from Bala's invention where initial splitting is only performed by said winning agent.

CLAIM LIMITATION F

Claim Limitation	Examiner's Position	Cited Language	Bala Position
forwarding split data index information resulting from said data splitting by said winning agent to said mediator	The combination of Kargupta in view of Kamath, disclose: forwarding split data index information (column 13, lines 64-66, Kamath; wherein when the list is being sorted this results in an data index) resulting from said data splitting by said winning agent to said mediator (column 14, lines 9-19, Kamath)	Kamath (column 14, lines 9-19) states: *Each processor evaluates each of the local feature lists to find the best local split (this is done in parallel by all processors).* *It communicates the local best splits and count statistics to all processors.* *Each processor determines the best global split (this is done in parallel by all processors).* *Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors.*	Again, Kamath does not disclose forwarding split data index information resulting from said data splitting by said winning agent. Instead, he discloses the forwarding of all split data index information by all the agents (or processors). This is a significant algorithmic difference.

CLAIM LIMITATION H

Claim Limitation	Examiner's Position	Cited Language	Bala Position
initiating data splitting by each of said plurality of agents other than said winning agent	The combination of Kargupta in view of Kamath, disclose: Initiating data splitting by each of said plurality of agents other than said winning agent (column 14, lines 17-26, Kamath)	Kamath (column 14, lines 17-26) states: *Split the Data. Each processor splits on the winning feature, and sends the ID numbers of its new left and right node data instances to all other processors.* *Then, each processor builds a hash table containing all the ID numbers, and information on which instances belong to which decision tree node.* *Next, each processor, for each feature, probes the hash table for each ID number to determine how to split that feature value.* *This process is carried out on the next unsolved decision tree.*	Again, Kamath does not disclose initiating data splitting by each of said plurality of agents "other than said winning agent". In fact, Kamath clearly states that data splitting is performed by all the agents (or processors).

CLAIM LIMITATION I

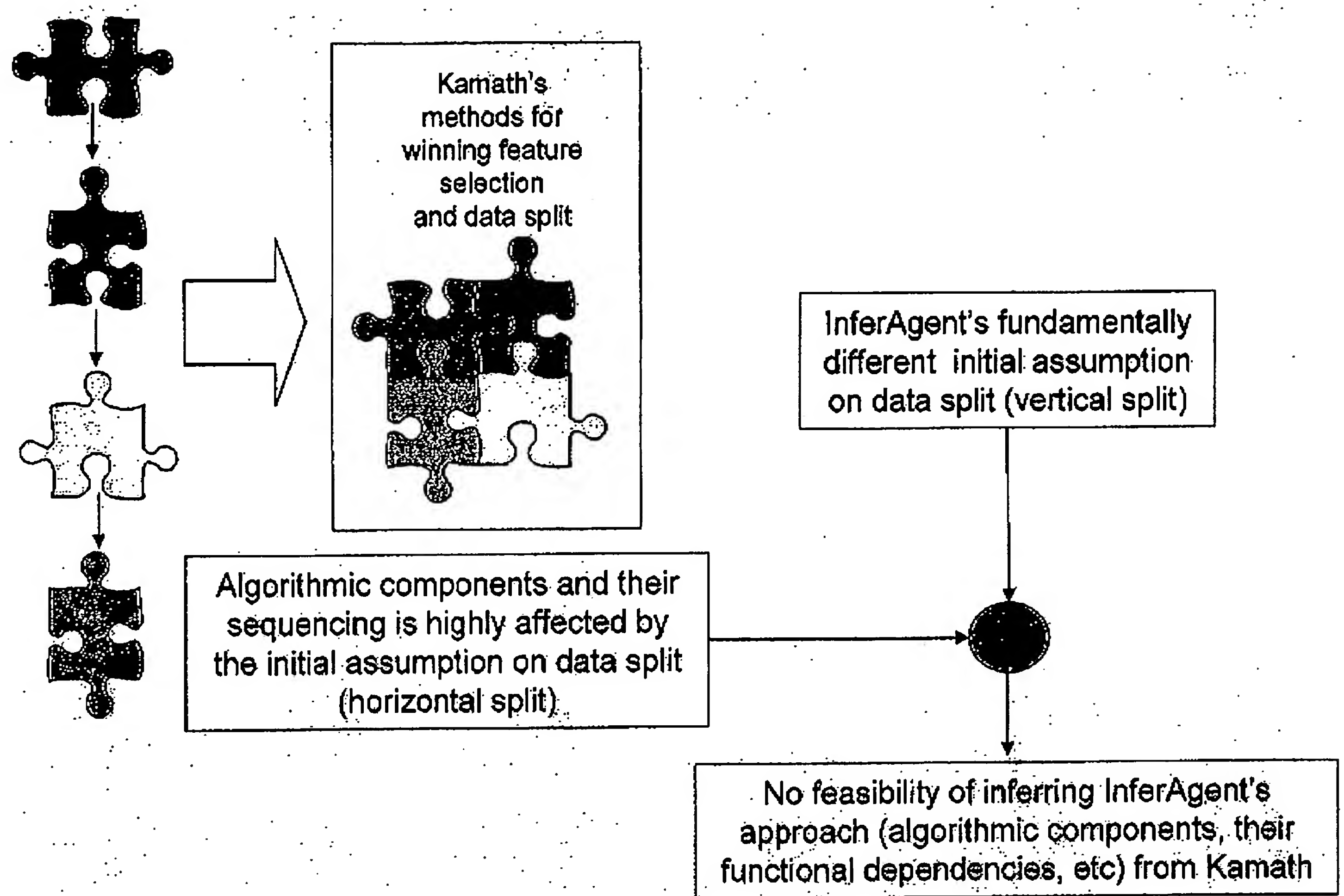
Claim Limitation	Examiner's Position	Cited Language	Bala Position
generating and saving partial rules	Cho discloses: generating and saving partial rules (pg. 2, lines 14-18, Cho)	Cho (pg. 2, lines 14-18) states: *Besides devising a new technique named 'fragmentation approach' in this paper, we also investigate various technical details, such as how large a data set in each fragment (local data set) is adequate, how many rules are to be generated in each fragment, and the corresponding selection criterion by which to choose the generated rules form the fragments to form a global rule set.*	Here, Cho clearly is not disclosing generating and saving partial rules. Instead, he is referring to the process of generating whole rules at each fragment and then using them to form a global rule set. In contrast, Bala's invention does not advocate constructing whole rules at each agent location. Instead, Bala's invention constructs rule "conditions" at each location. For example consider the following two rules: A & B -> C1 (Rule 1) C & D -> C3 (Rule 2) Cho suggests generating all of Rule 1 at one fragment and all of Rule 2 at another and then combining them to get the rule set. Instead, Bala's invention would generate A at one location, and B at another location, so that combining them through a mediator would finally reveal the whole rule: A & B -> C1

CLAIM LIMITATION J

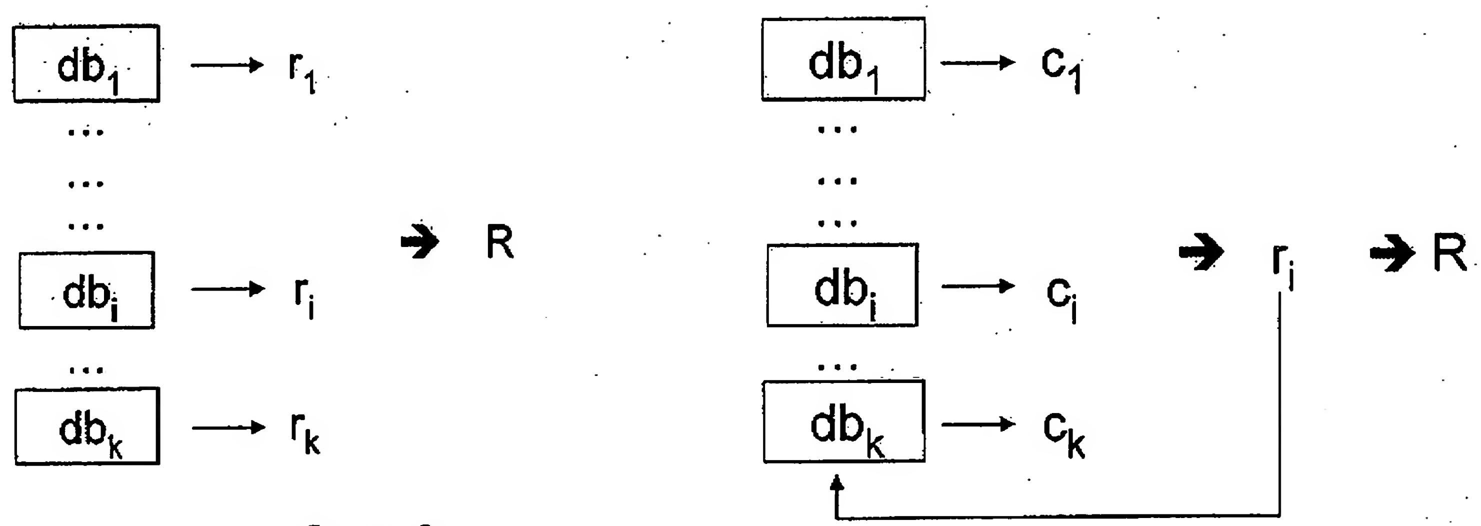
Claim Limitation	Examiner's Position	Cited Language	Bala Position
outputting complete rules to said plurality of agents	Cho discloses: outputting complete rules to said plurality of agents (pg. 4, lines 24-25, Cho)	Cho (pg. 4, lines 24-25) states: *Another strategy is to loosely account on an number of different inductive learning algorithms by integrating their collective output concepts.*	Here, Cho is by no means disclosing outputting complete rules to said plurality of agents. Instead, he is referring to the integration of the outputs of a number of different learning algorithms. This deals with the topic of "multiple classifier" approaches, which is a whole discipline unto itself and does not have anything to do with Bala's invention.

Backup

Feasibility of Inferring



Cho's Algorithm



Case 3

Repetitive process of assembling rules globally by contributing conditions each distributes locales